Kruskall Wallis

Professor: Pedro M. Almeida-Junior

28 de maio de 2021

Departamento de Estatística (UEPB)

Kruskal-Wallis

- A análise de variância de um fator de Kruskal-Wallis por postos é um teste extremamente útil para decidir se k amostras independentes provêm de populações diferentes

- Valores amostrais quase sempre diferem um pouco, e a questão é se as diferenças entre as amostras significam genuínas diferenças entre as populações ou se elas representam meramente o tipo de variações que seriam esperadas entre amostras aleatórias de uma mesma população.

- A técnica de Kruskal-Wallis testa a hipótese nula de que as k amostras provêm da mesma população ou de populações idênticas com a mesma mediana.

Hipóteses do teste de Kruskall-Wallis

- Podemos escrever a **hipótese nula de que as medianas são as mesmas** como $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k$; e a hipótese alternativa pode ser escrita como $H_1 : \theta_i \neq \theta_j$ para alguns grupos i e j. Isto é, se a hipótese alternativa é verdadeira, pelo menos um par de grupos tem medianas diferentes.

- Sob a hipótese nula, o teste admite que as variáveis sob estudo têm a mesma distribuição contínua subjacente; assim, ele **requer pelo menos mensuração ordinal da variável**.

- Ao aplicar a análise de variância de um fator de Kruskal-Wallis por postos, os dados são colocados em uma tabela de duas entradas com cada coluna representando cada amostra ou grupo sucessivo.

Amostras					
1	2	...	j	...	k
X_{11}	X_{12}	...	X_{1j}	...	X_{lk}
X_{21}	X_{22}	...	X_{2j}	...	X_{2k}
...	...	...	...	...	...
X_{i1}	X_{i2}	...	X_{ij}	...	X_{ik}
...	...	...	...	...	...
$X_{n_1 1}$	$X_{n_2 2}$	...	$X_{n_i j}$	...	$X_{n_k k}$
n_1	n_2	...	n_i	...	n_k

- No cálculo do teste de Kruskal-Wallis, cada uma das N observações é substituída por postos. Isto é, todos os escores de todas as k amostras são colocados juntos e organizados através de postos em uma única série.

- O menor escore é substituído pelo posto 1 , o seguinte menor escore é substituído pelo posto 2 e o maior escore é substituído pelo posto N, onde N é o número total de observações independentes nas k amostras.

- O teste de Kruskal-Wallis trabalha com as diferenças entre os postos médios para determinar se elas são tão discrepantes que provavelmente não tenham vindo de amostras que tenham sido extraídas de uma mesma população.

- Daremos duas formas para o teste de Kruskal-Wallis e os termos necessários para calcular a estatística de Kruskal-Wallis:

$$KW = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(\bar{R}_j - \bar{R}\right)^2$$

ou

$$KW = \left[\frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \bar{R}_j^2\right] - 3(N+1)$$

em que

$k =$ número de amostras ou grupos $n_j =$ número de casos na j-ésima amostra

$N =$ número de casos na amostra combinada (a soma dos $n_j' s$)

$R_j =$ soma dos postos na j -ésima amostra ou grupo

$\bar{R}_j =$ média dos postos na j -ésima amostra ou grupo

$\bar{R} = (N+1)/2 =$ média dos postos na amostra combinada

- Se H_0 é verdadeira, então a distribuição amostral da estatística KW pode ser calculada e a probabilidade de observar diferentes valores de KW pode ser tabelada.

- Quando há mais do que $k = 3$ grupos e quando o número de observações em cada grupo excede cinco, a distribuição amostral de KW é bem aproximada pela distribuição χ^2 com $gl = k - 1$.

- A aproximação torna-se melhor quando ambos, o número de grupos, k, e o número de observações dentro de cada grupo, n_j, crescem.

Exemplo

Exemplo: Em um estudo sobre os efeitos pulmonares em porquinhos-da-índia, sensibilizados com ovalbumina (OA) ao ar normal, benzaldeído ou acetaldeído. No final da exposição, os porquinhos-da-índia foram anestesiados e as respostas alérgicas foram avaliadas no lavado broncoalveolar (LBA). Uma das variáveis de resultado examinadas foi a contagem de células eosinófilas, um tipo de glóbulo branco que pode aumentar com alergias. A Tabela seguinte fornece a contagem de células de eosinófilos $\left(\times 10^6\right)$ para os três grupos de tratamento.

Tabela 1: Número de células alveolares $(\times 10^6)$

Ar Normal	benzaldeído	acetaldeído
0.55	0.81	0.65
0.48	0.56	13.69
7.8	1.11	17.11
8.72	0.74	7.43
0.65	0.77	5.48
1.51	0.83	0.99
0.55	0.81	0.65

Comparações múltiplas entre tratamentos

- Quando o valor obtido de KW é significante, ele indica que pelo menos um dos grupos é diferente de pelo menos um dos outros grupos. Mas ele não informa ao pesquisador quais deles são diferentes, nem quantos grupos são diferentes de cada um dos outros.

- O que é necessário é um procedimento que torne possível determinar quais grupos são diferentes. Isto é, gostaríamos de testar a hipótese $H_0 : \theta_u = \theta_v$ contra a hipótese $H_1 : \theta_u \neq \theta_v$ para alguns grupos u e v.

- Existe um procedimento simples para determinar quais pares de grupos são diferentes. Começamos obtendo as diferenças $\left|\bar{R}_u - \bar{R}_v\right|$ para todos pares de grupos.

- Quando o tamanho da amostra é grande, estas diferenças têm distribuição aproximadamente normal.

- Vamos atribuir uma diferença significativa dos pares individuais de diferenças quando a seguinte desigualdade for satisfeita:

$$\left|\bar{R}_u - \bar{R}_v\right| \geq z_{\alpha/k(k-1)} \sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_u} + \frac{1}{n_v}\right)} \tag{1}$$